

# Spectro-Temporal Processing of Dynamic Broadband Sounds In Auditory Cortex

Shihab Shamma  
Jonathan Simon\*  
Didier Depireux  
David Klein

Institute for Systems Research  
& Department of Electrical Engineering  
University of Maryland

Supported in part by  
Office of Naval Research MURI # N00014- 97- 1- 0501  
National Institute on Deafness and Communication Disorders # T32 DC00046-01  
and National Science Foundation # NSFD CD8803012

These slides are available at <<http://www.isr.umd.edu/CAAR/pubs.html>>

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>1999</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-1999 to 00-00-1999</b>	
4. TITLE AND SUBTITLE <b>Spectro-Temporal Processing of Dynamic Broadband Sounds in Auditory Cortex</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>University of Maryland, Department of Electrical Engineering , Institute for Systems Research, College Park, MD, 20742</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>29</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

# Outline

- **Introductory Material**
- **Stimulus Types**
- **Spectro-Temporal Response Fields**
- **Predicting Responses to Novel Stimuli**
- **Non-Linearities**

- **Introductory Material**
- Stimulus Types
- Spectro-Temporal Response Fields
- Predicting Responses to Novel Stimuli
- Non-Linearities

# Motivation

- **The Big Quest**

Teasing out “function” of Primary Auditory Cortex (AI)

function  $\approx$   $\begin{cases} \text{which sounds/features evoke responses?} \\ \text{how are they encoded into spike trains?} \end{cases}$

- **Broadband and dynamic sounds**

- Evoke strong, sustained, dynamic responses in AI
- Many natural sounds  
e.g. speech, vocalizations, backgrounds

- **Reasonable quest**

Quantitative measure of how spikes encode sound features

- Quantitative descriptor
- Quantitative predictor
- Visual tool—*a la* Spectral Response Field/Tuning Curve

- **Compromise** from quantitative necessity

- Restrict broadband and dynamic sounds to mathematically simple subset:
- **Noise**—strongly modulated in **spectrum** and **time**
- not a severe compromise

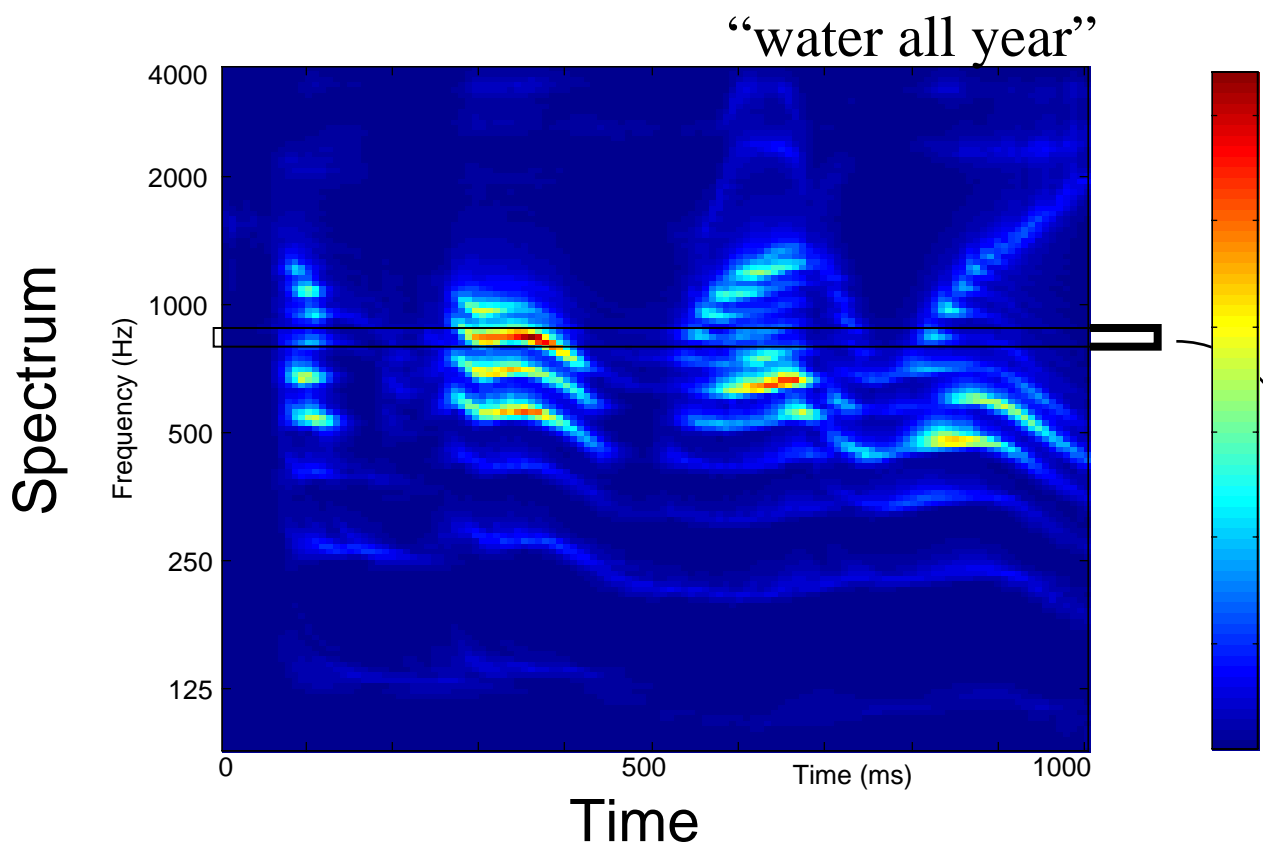
- **Spectro-Temporal Receptive Field (STRF)** meets criteria:

- Quantitative descriptor
- Quantitative predictor
- Visual Tool

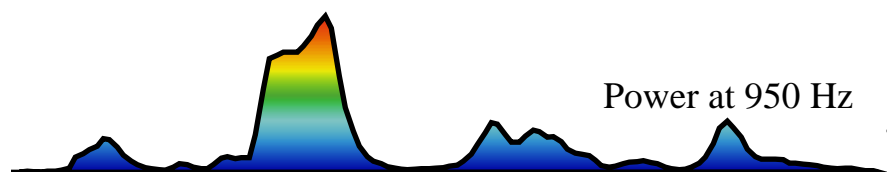
# Sound Features

- **Spectro-Temporal Features of Any Sound**
- Spectral content of sound as a function of time.

Which spectral frequency bands have enhanced power?  
Which spectral frequency bands have diminished power?  
How do these change as a function of time?



- $\log(f) \approx$  linear cochlear distance

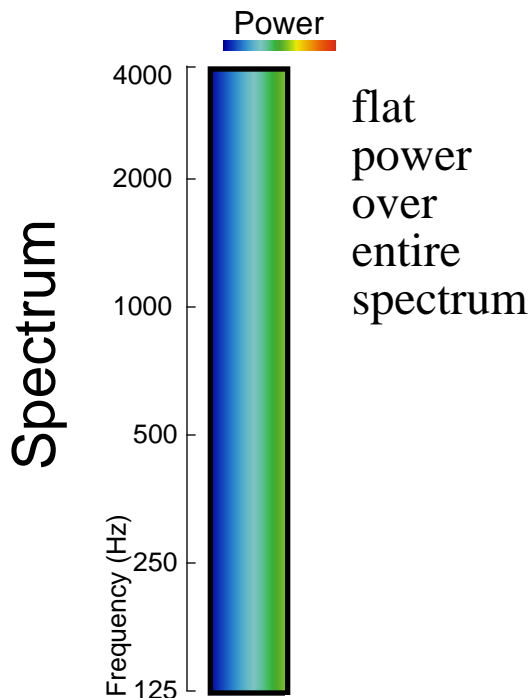


- Characterization from cross-section is limited

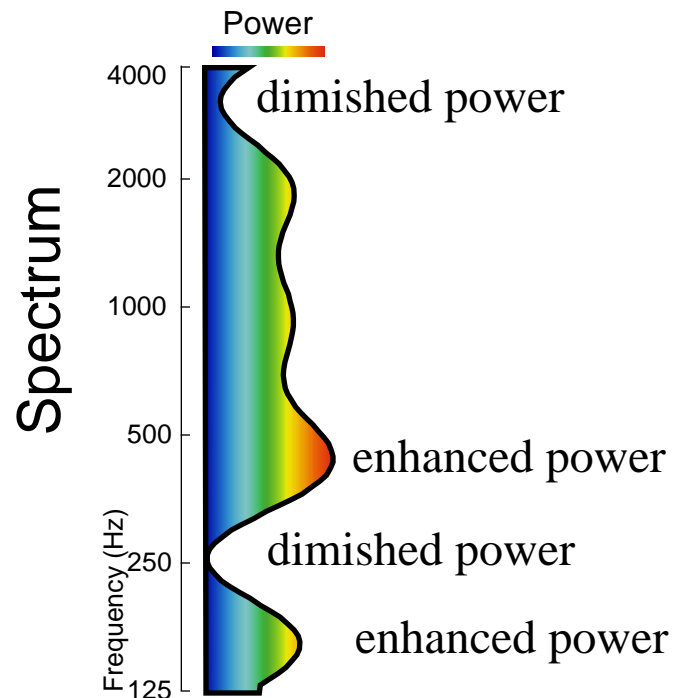
# Stimulus Construction

- Pink Noise = flat power density in octaves [ $\log(f)$ ] not white

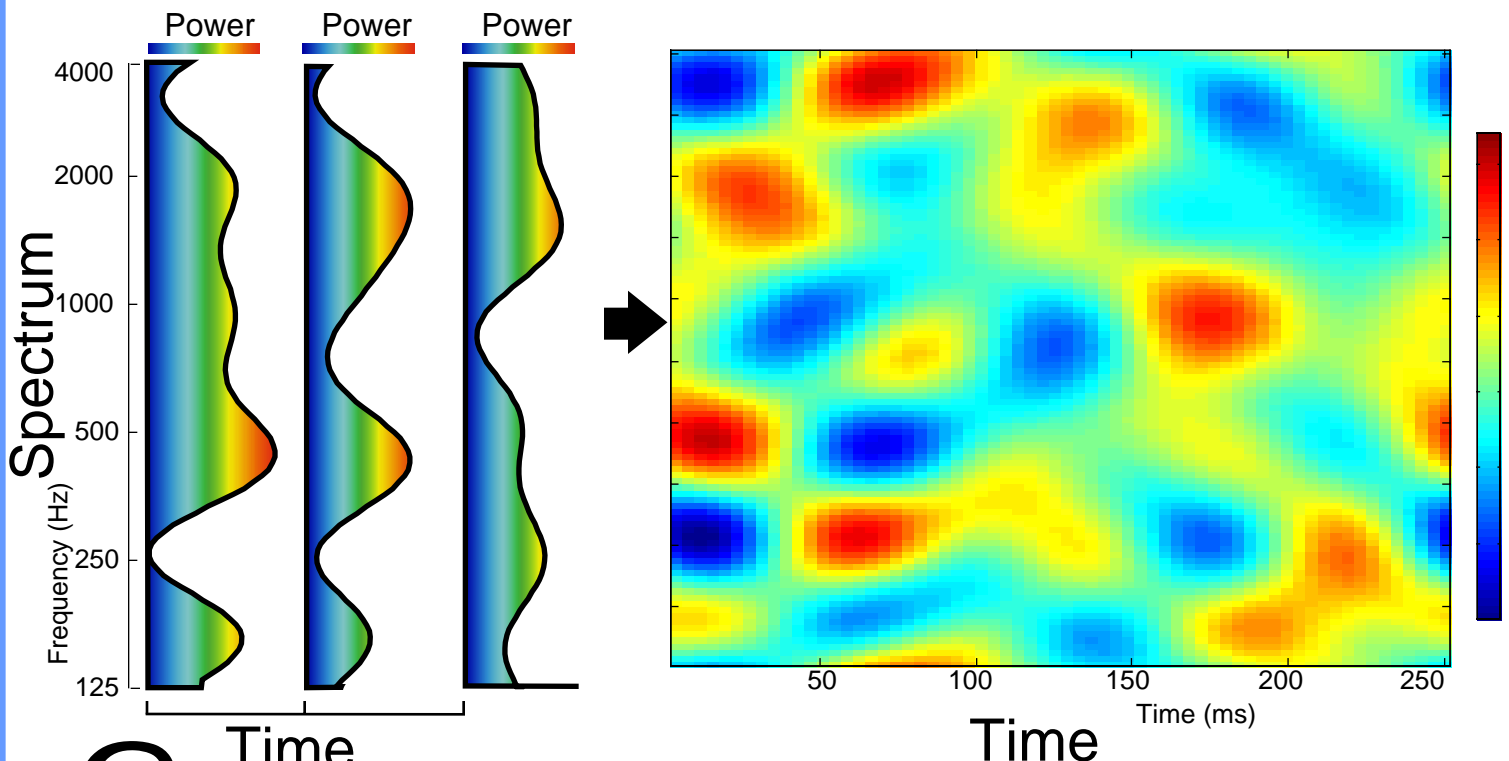
- *Unmodulated noise (flat)*



- *Spectrally modulated noise*

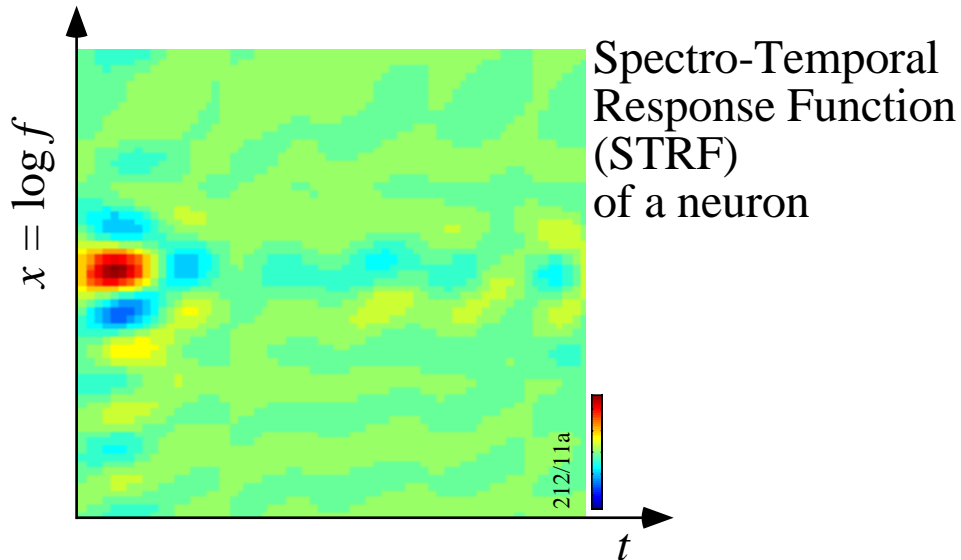


- *Spectro-Temporally modulated noise*



# Spectro-Temporal Response Field (STRF)





- Spectro-temporal response field of neuron is the standard response field, made time-dependent.
- Frequencies mapped along cochlea on log frequency axis



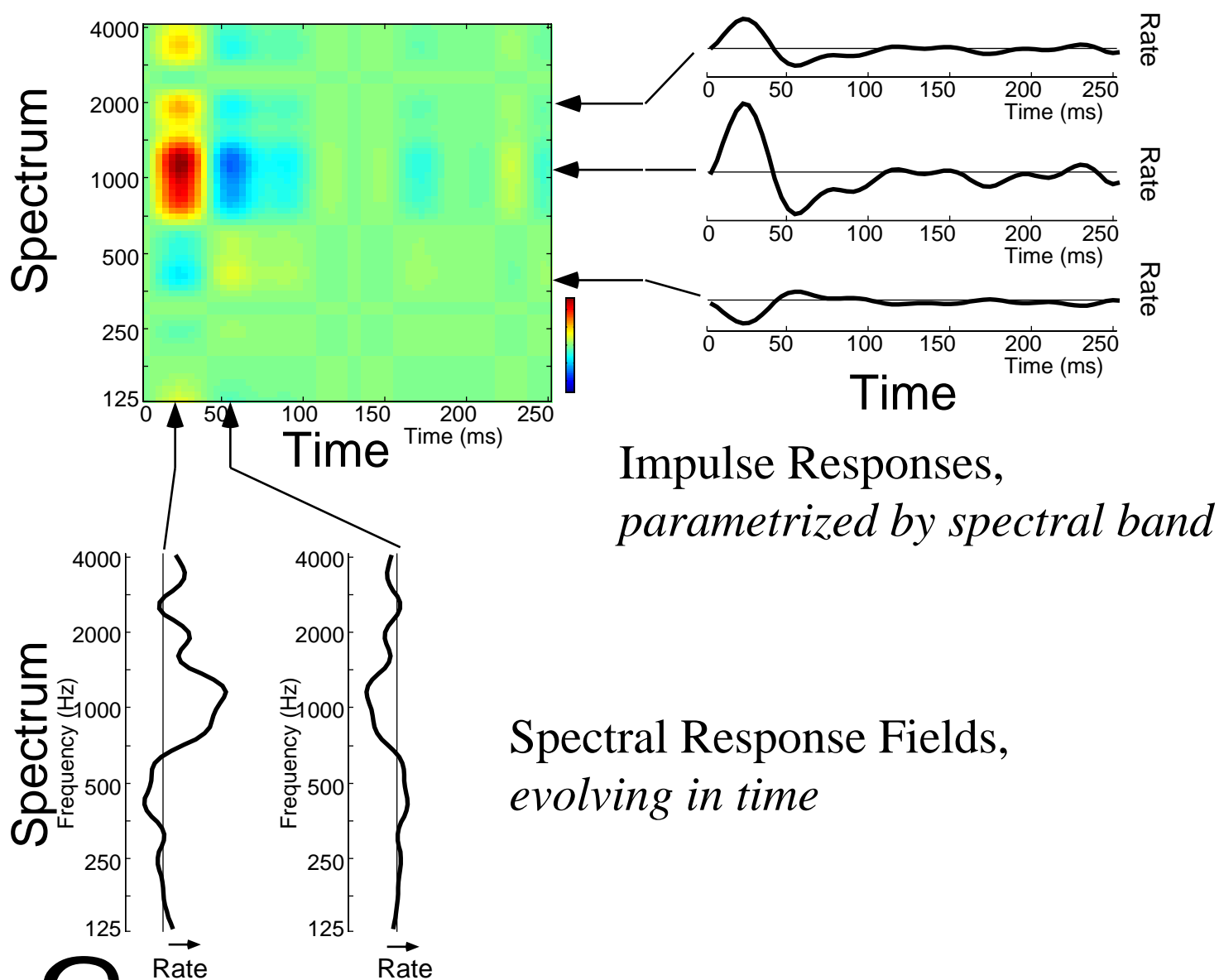


# Interpreting STRFs

## Stimulus Effect on Rate

STRF region	Stimulus Power	Spike rate contribution
Excitatory ●	Enhanced ●	Faster 
Inhibitory ●	Enhanced ●	Slower 
Excitatory ●	Diminished ●	Slower 
Inhibitory ●	Diminished ●	Faster (!) 

## Cross-section interpretations



# Spectro-Temporal Response Field (STRF) & its Transfer Function

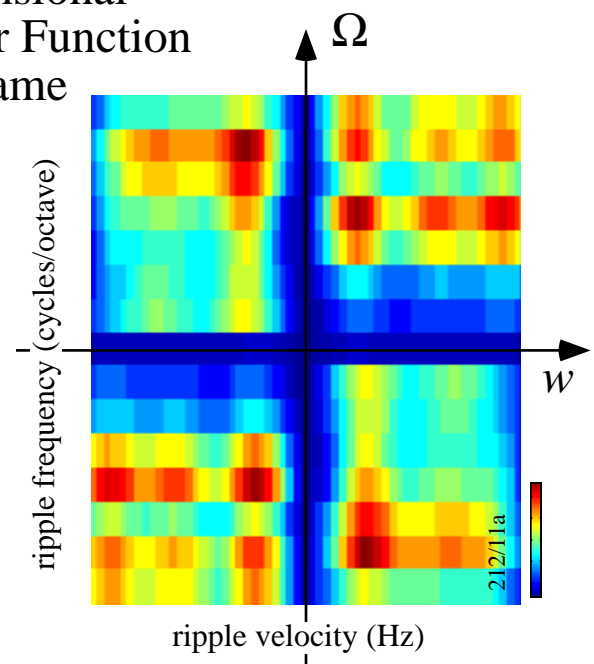
2 Dimensional Fourier Transform

$$\int [.] \exp(\pm 2\pi j \Omega x \pm 2\pi j \omega t)$$

Inverse Transform

- Its Fourier transform is the transfer function.
- Analysis is often conceptually simpler in the Fourier domain.
- Either can be used to predict the linear response to *any* broadband dynamic sound.

2 Dimensional  
Transfer Function  
of the same  
neuron



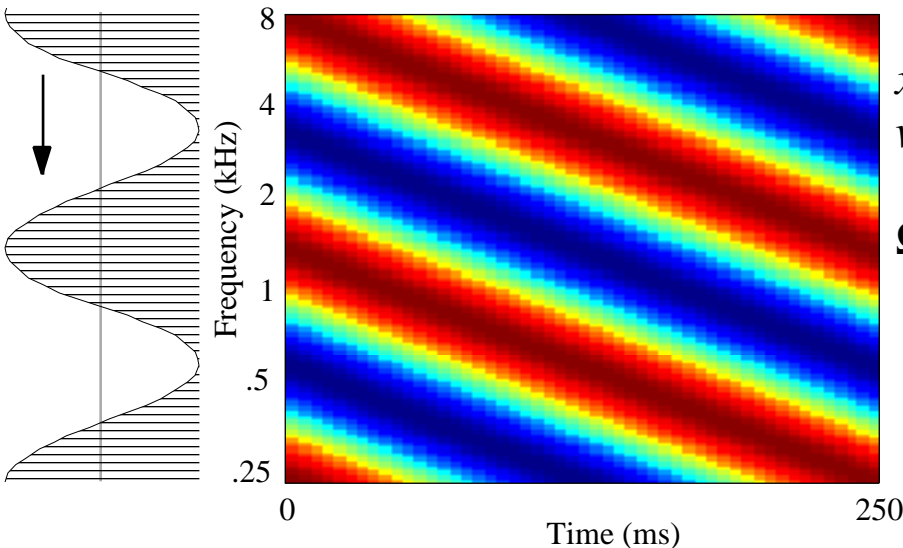
- Introductory Material
- **Stimulus Types**
- Spectro-Temporal Response Fields
- Predicting Responses to Novel Stimuli
- Non-Linearities

# Single Moving Ripple

## Single Ripple in Spectro-Temporal Space (*Spectrogram*)

Ripples are auditory “gratings” whose spectral envelope is a sinusoid along the  $\log(f)$  axis. At any time  $t$  and any frequency  $x$ , the amplitude  $S(t,x)$  is given by:

$$S(t,x) = \sin[2\pi w t + 2\pi \Omega x + \phi]$$



$$x = \log_2[f / f_0]$$

$w$  = ripple velocity,

e.g. 4 Hz = 4 cycles/s

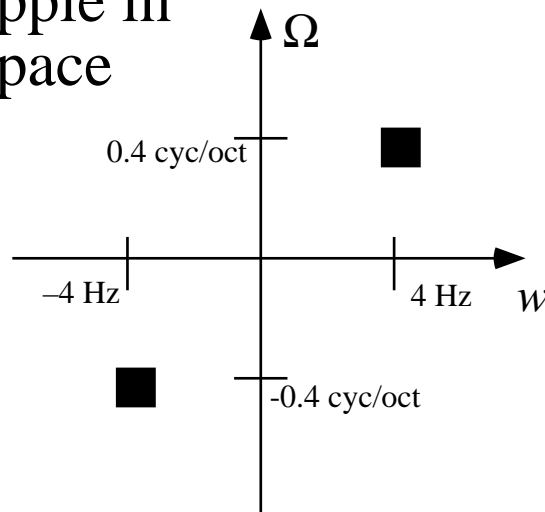
$\Omega$  = ripple density,

e.g. 0.4 cycles/octave

= 2 cycles/5 octaves

$$\int [.] \exp(\pm 2\pi j \Omega x \pm 2\pi j w t)$$

## Single Ripple in Fourier Space



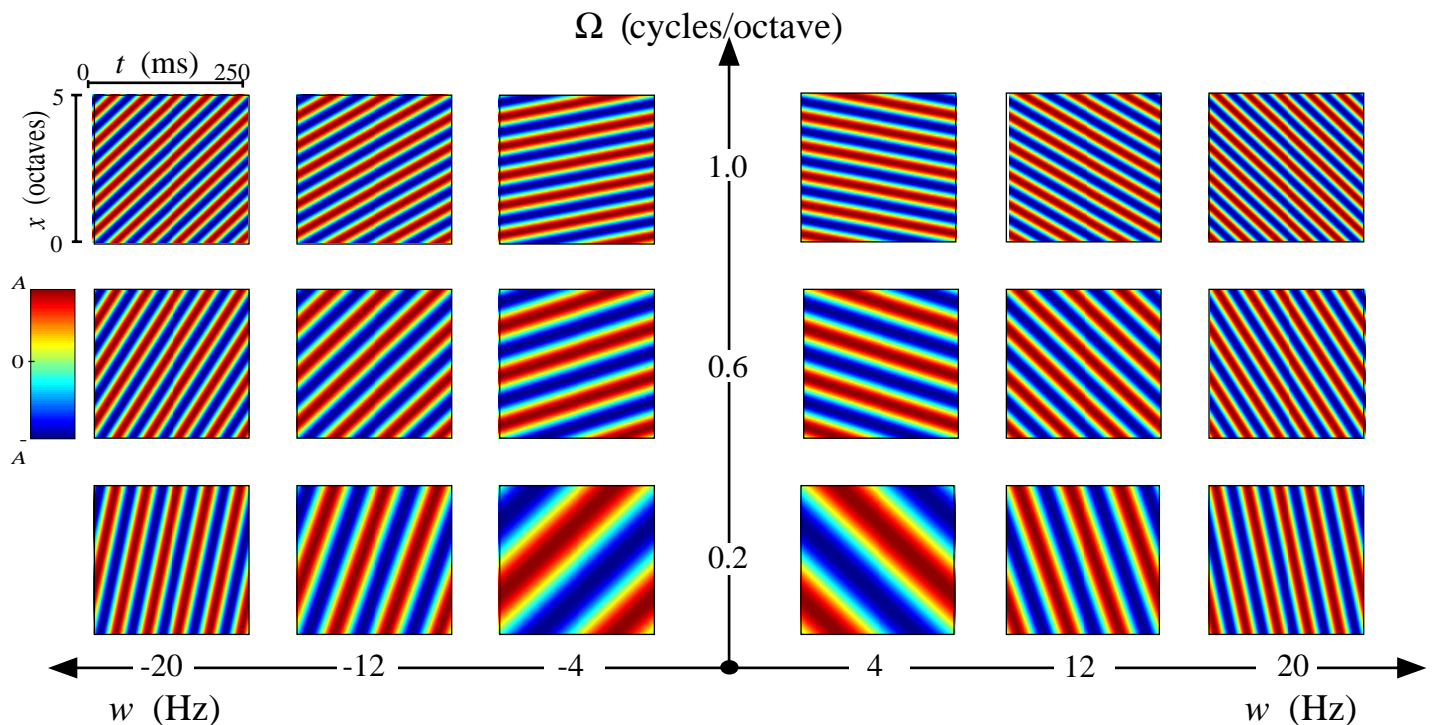
The Fourier transform of a single moving sinusoid has support only on a single point (and its complex conjugate).

# Multiple Individual Ripples

$$S(t,x) = \sin[2\pi w t + 2\pi \Omega x + \phi]$$

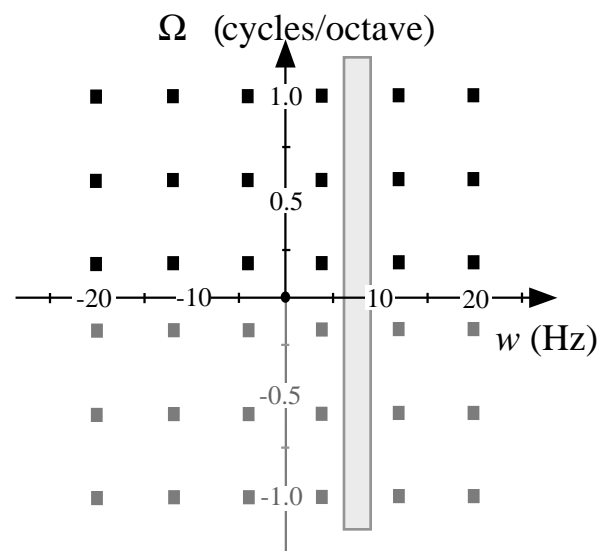
$$x = \log_2[f/f_0]$$

$w$  = ripple velocity  
 $\Omega$  = ripple density

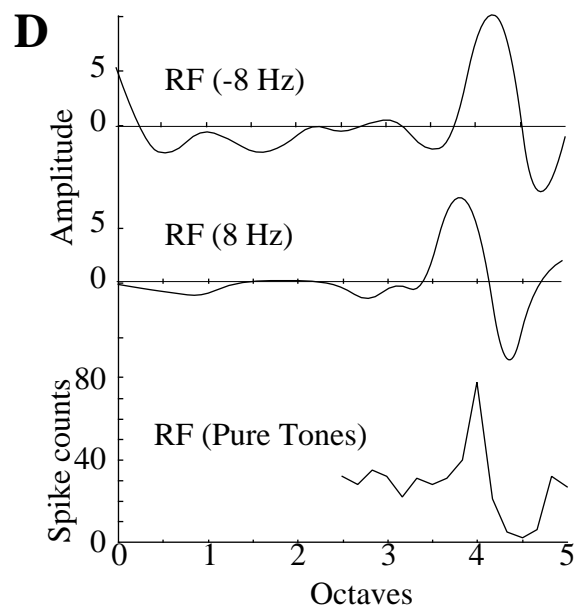
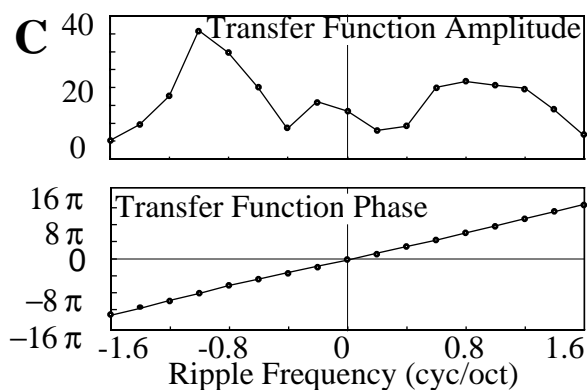
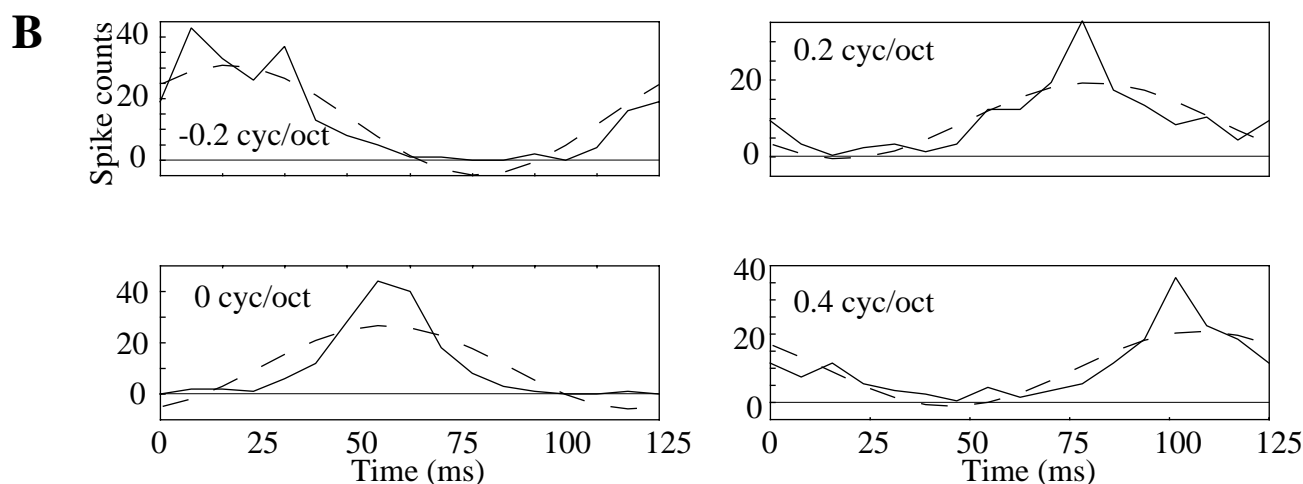
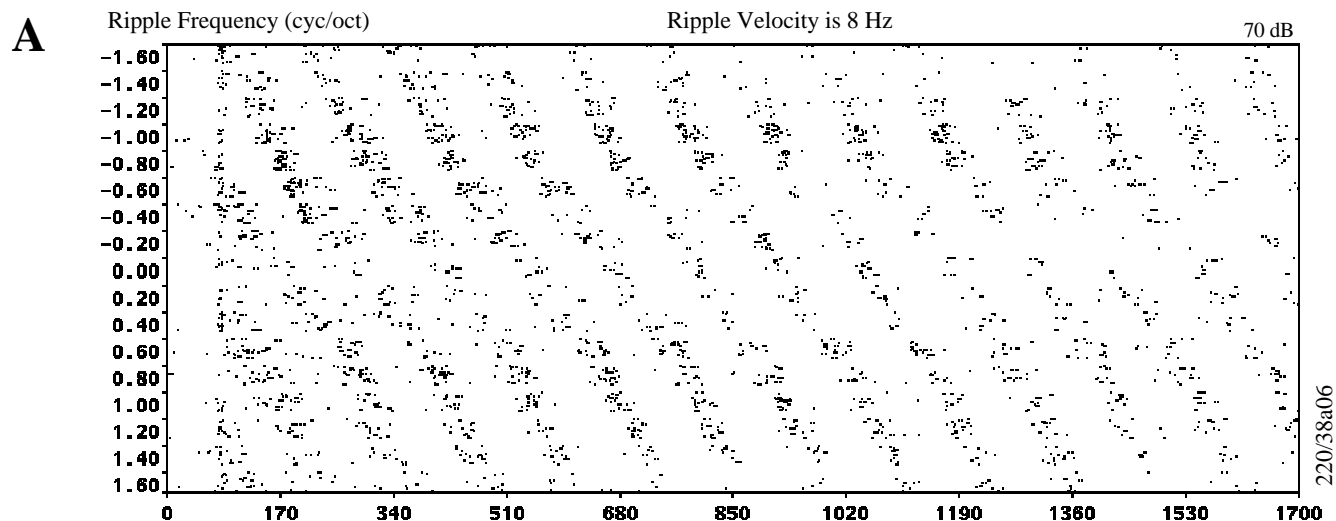


The Fourier space of the spectrograms. We probe a cell at different velocities  $w$  and different densities  $\Omega$ , and quantify the response for up and down-moving sounds.

Any ripple in the lower half-plane is equivalent to a ripple in the upper-half plane.



# Spike Train Measurements

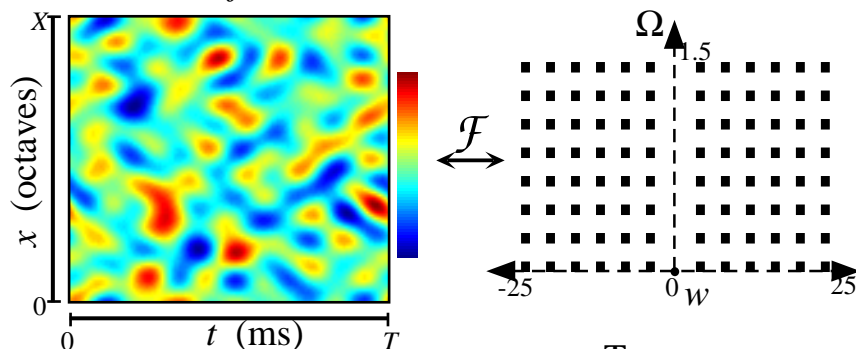


Spike events in (A) are turned into period histograms in (B). The amplitudes and phases give the transfer function in (C), which can be inverse Fourier transformed to give Response Fields in (D).

# Spectro-Temporal Noise

To speed up the characterization of a cell's response, we use combinations of ripples of *all* velocities  $w$  and densities  $\Omega$ , with random phases.

$$S^{\text{noise}}(t, x) = \sum_j \sum_k \sin[2\pi w_j t + 2\pi \Omega_k x + \phi_{j,k}]$$

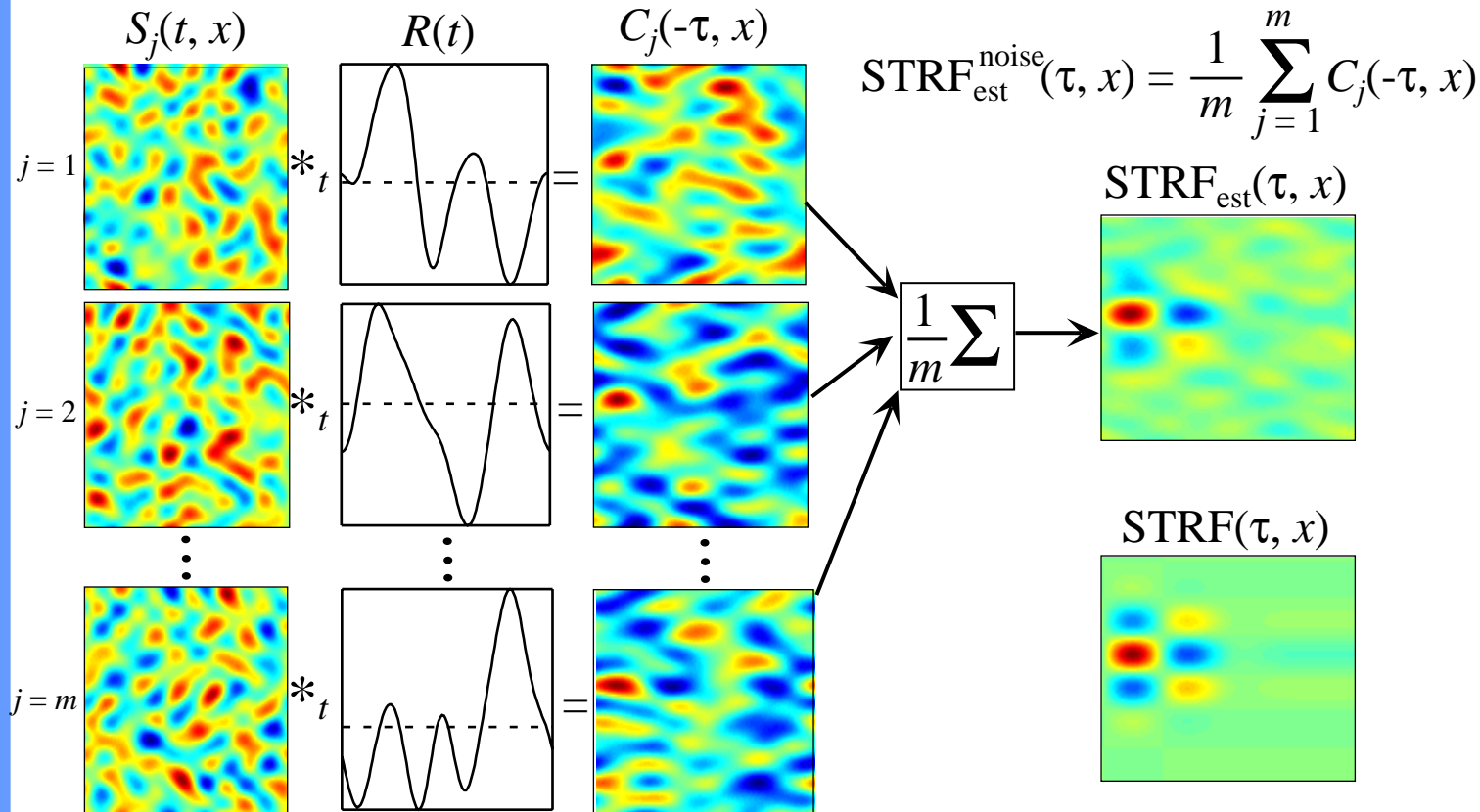


Spectro-Temporal  
generalization of  
white noise

$$\text{Cross-Correlation: } C(\tau, x) = \frac{1}{T} \int_0^T S(t, x) R(t-\tau) dt = \frac{1}{T} \sum_k S(t_k - \tau, x)$$

= Spike-Triggered  
Average

- $C(\tau, x)$  contains cross terms
- Cross terms have random phase and can be attenuated by averaging over multiple, random-phase stimuli.



Simulation

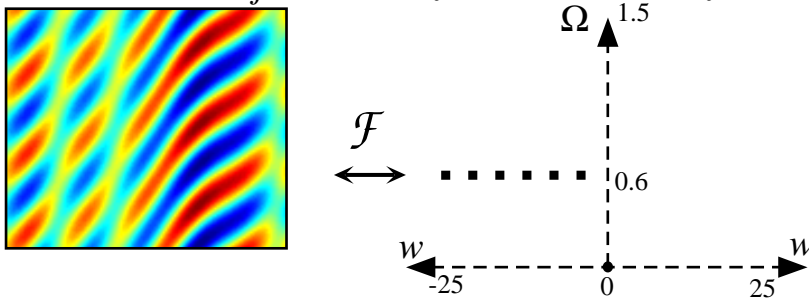


# Temporally Orthogonal Ripple Combinations (TORCs)

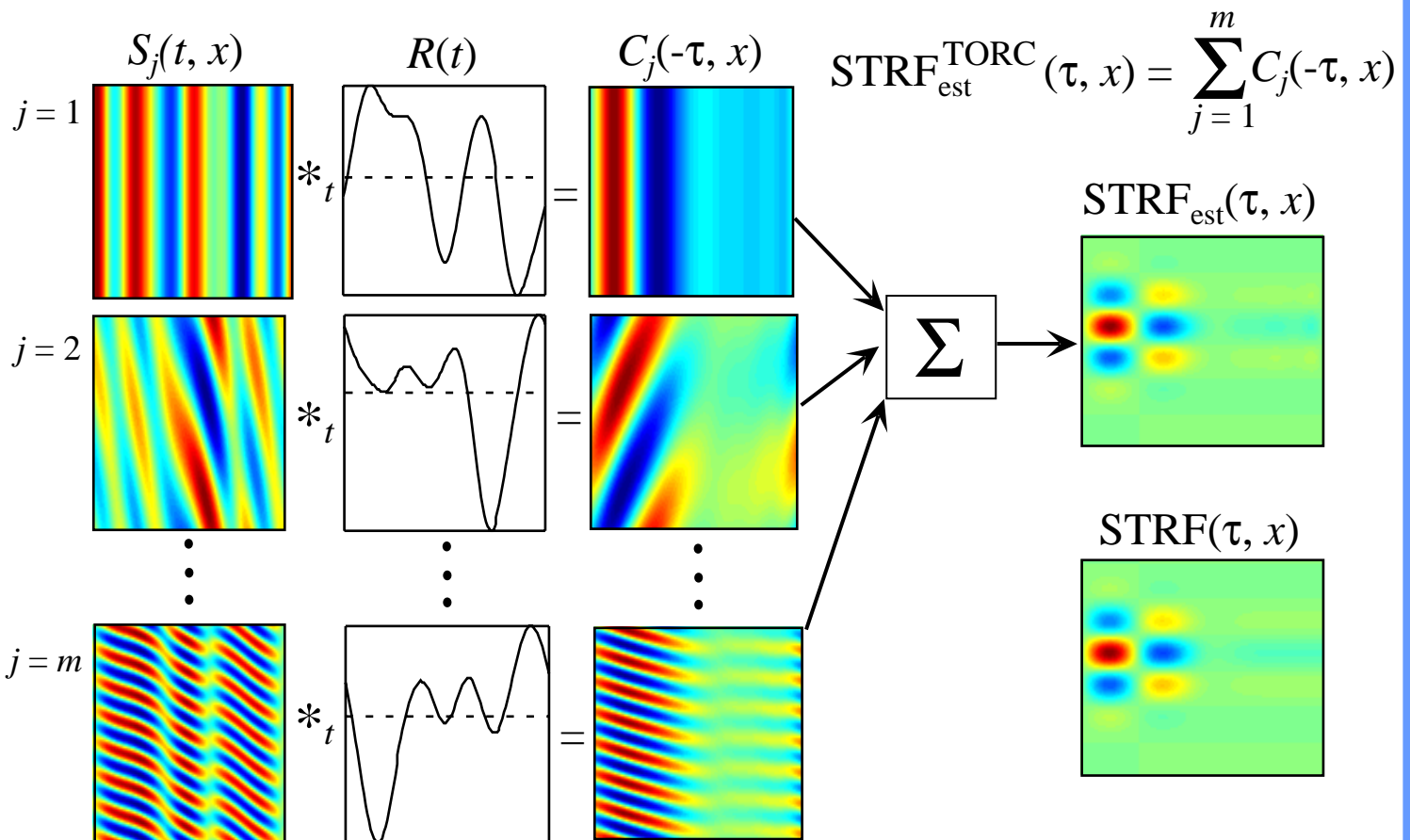
To eliminate interference from cross-terms, we use specific combinations of ripples with differing velocities  $w$  and random phases.

- Stimuli are composed only of ripples with different ripple velocities.
- Each stimulus contains ripples which cover the same range of ripple velocities, but at different ripple frequencies.
- Multiple stimuli are still needed to present a complete set of ripples.

$$S^{\text{TORC}}(t, x) = \sum_j \sin[2\pi w_j t + 2\pi \Omega_k x + \phi_{j,k}]$$



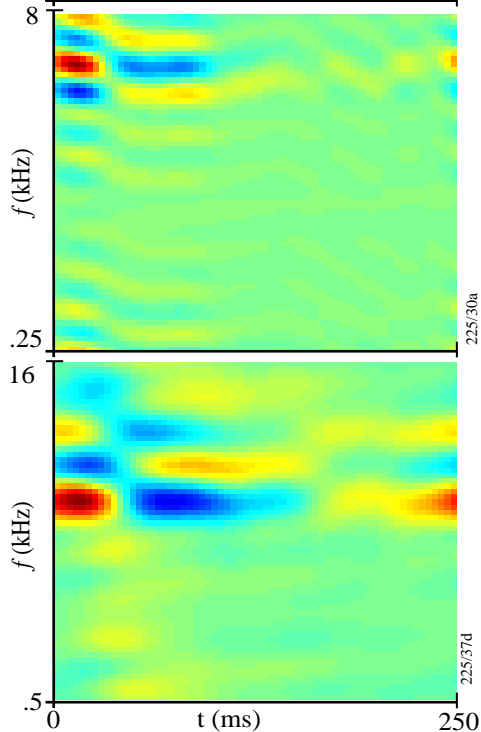
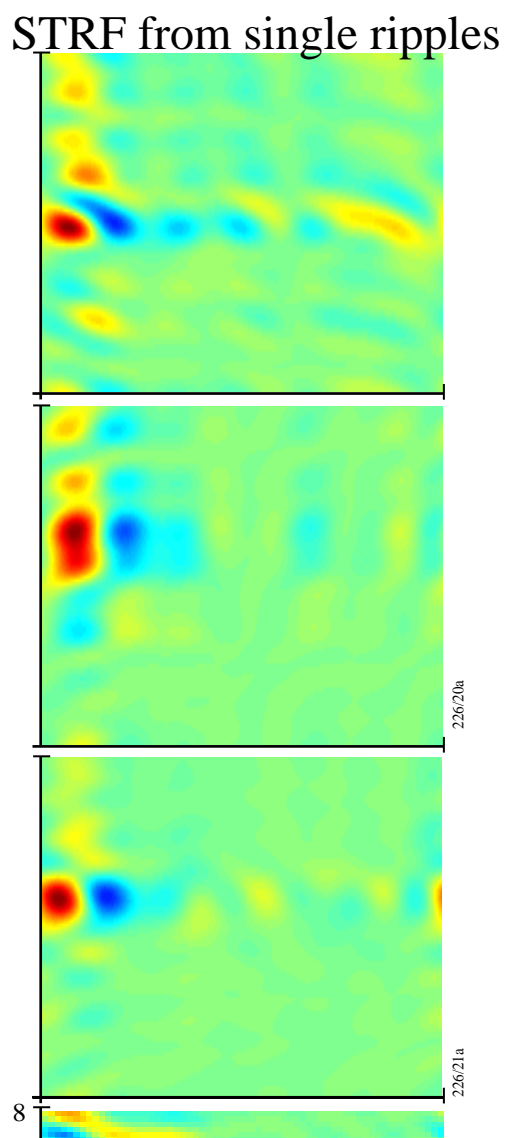
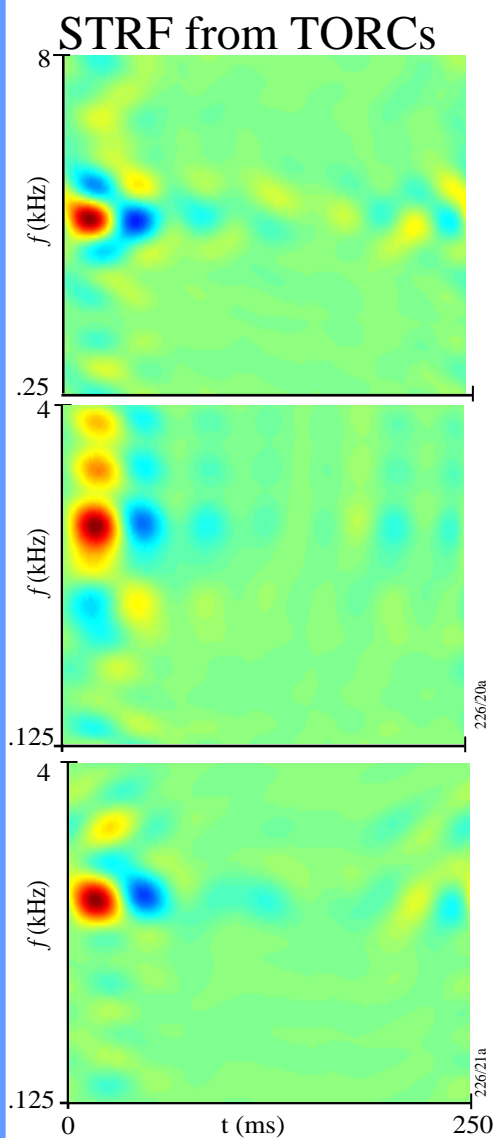
TORCs are better suited for temporal cross-correlation because there are no cross terms. The resulting estimates are robust, use short-duration stimuli, and are quickly computed.



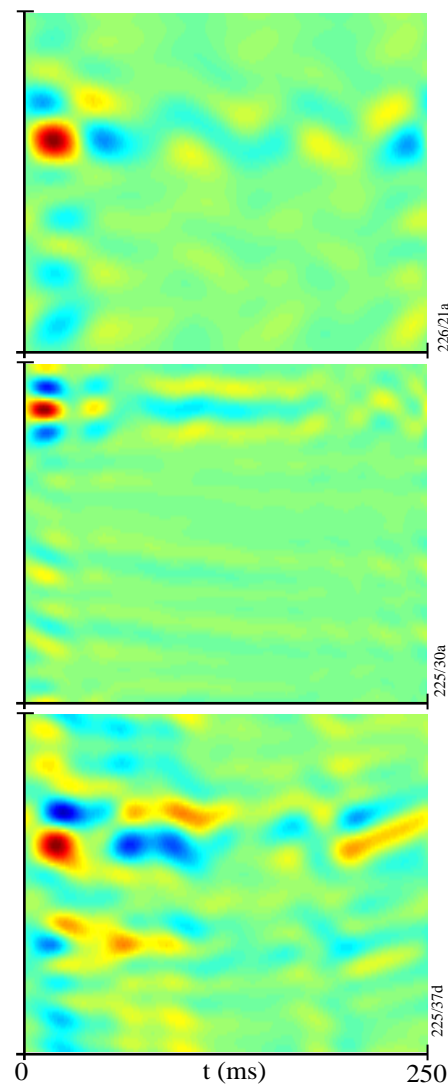


- Introductory Material
- Stimulus Types
- **Spectro-Temporal Response Fields**
- Predicting Responses to Novel Stimuli
- Non-Linearities

# STRFs Compared

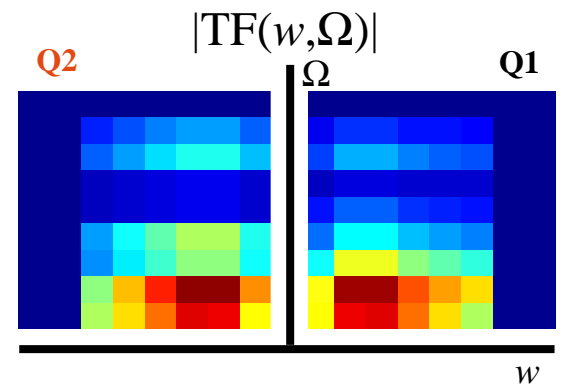
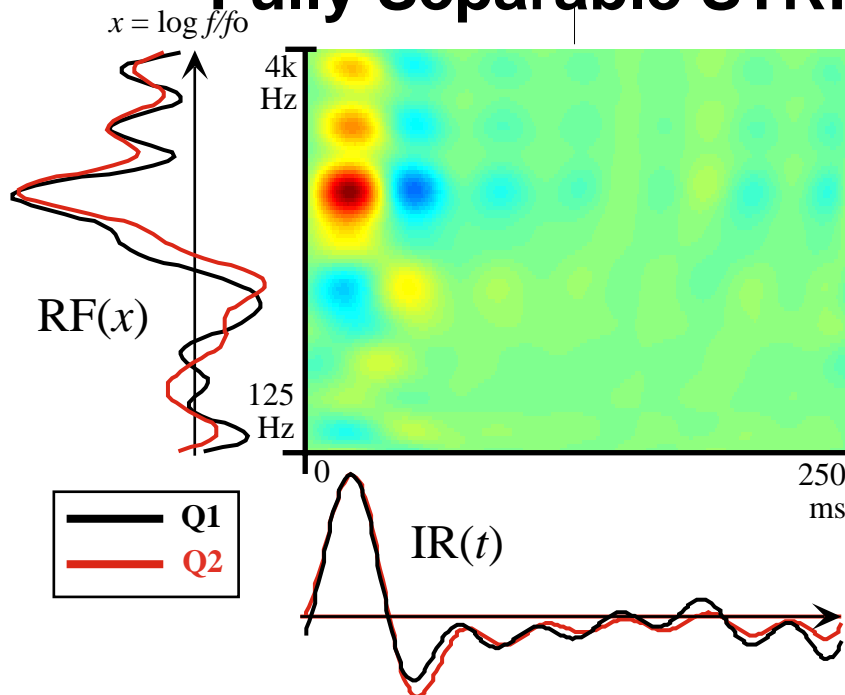


STRF from Spectro-Temporal Noise



# Separability

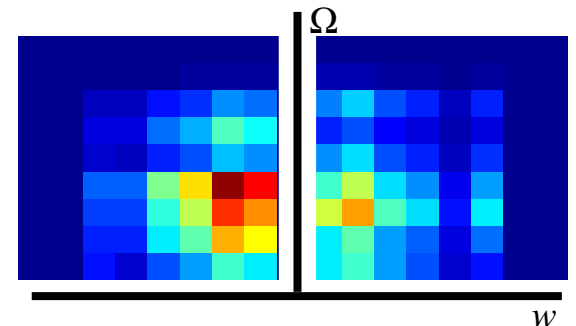
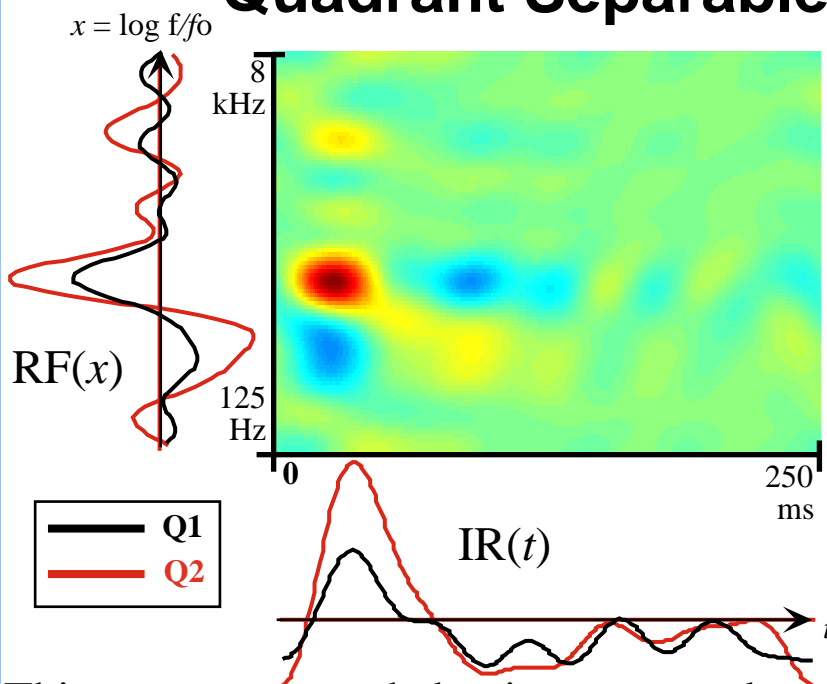
## Fully Separable STRF



The STRF and TF are a product of a single spectral response function with a single temporal response function.

Shown above are the impulse responses (IR) and receptive fields (RF) derived from quadrant 1 (black) and quadrant 2 (red) of the transfer function by inverse Fourier transformation.

## Quadrant Separable STRF



The STRF is not separable, but each quadrant of the transfer function is, i.e., there are different spectral and temporal responses for upwards and downwards frequency modulation.

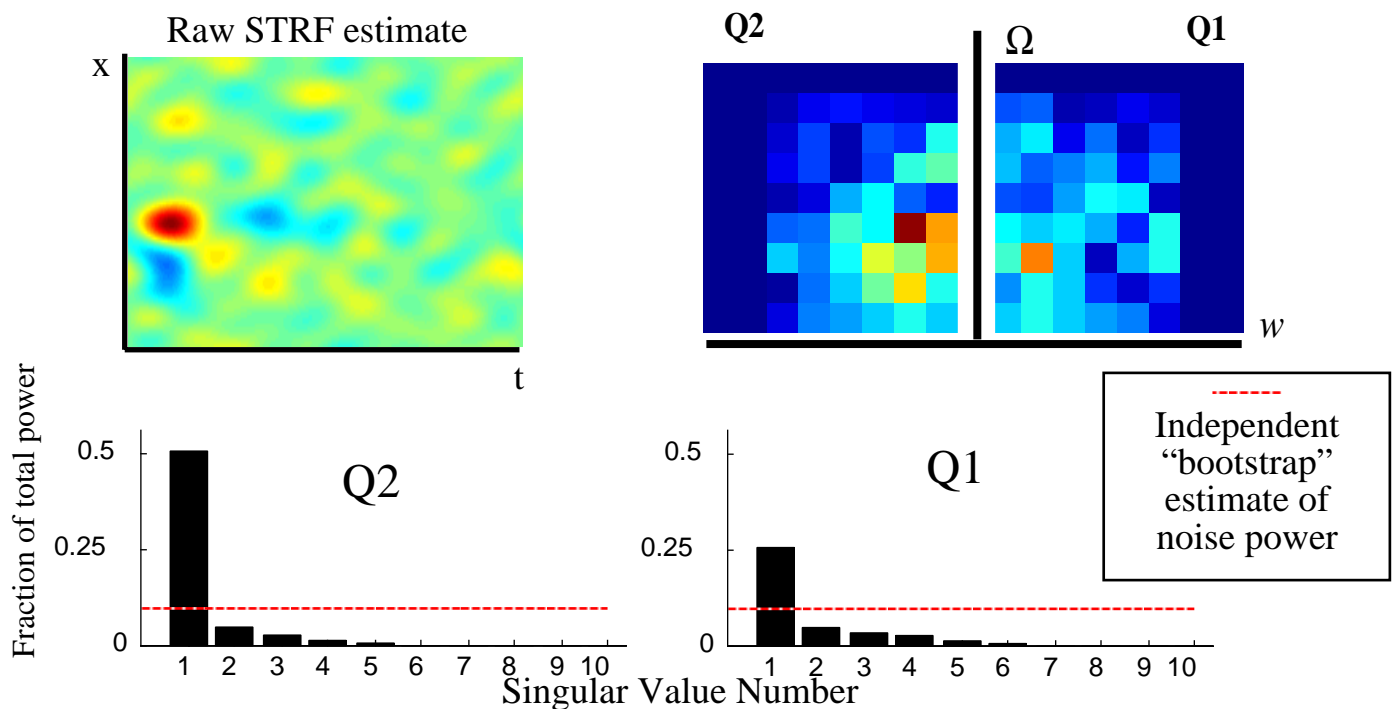
This neuron responded twice as strongly to rising frequencies as it did to falling frequencies.

# Singular Value Decomposition (SVD)

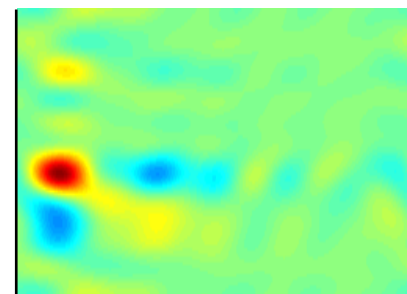
- Evaluation of Separability
- Noise Reduction

Singular Value Decomposition (SVD) decomposes the matrix into a sum of separable matrices, ordered by their overall magnitudes. The first  $k$  components sum to a matrix which minimizes the power of the remaining components.

We apply SVD to each quadrant of the transfer function.



SVD naturally separates the signal and noise components of a matrix. Typically, large jumps in the singular values indicate where the separation occurs. Noise is removed by discarding the lower-magnitude components.



**Without prior assumptions, SVD indicates that a large majority of STRFs in AI are quadrant separable.**

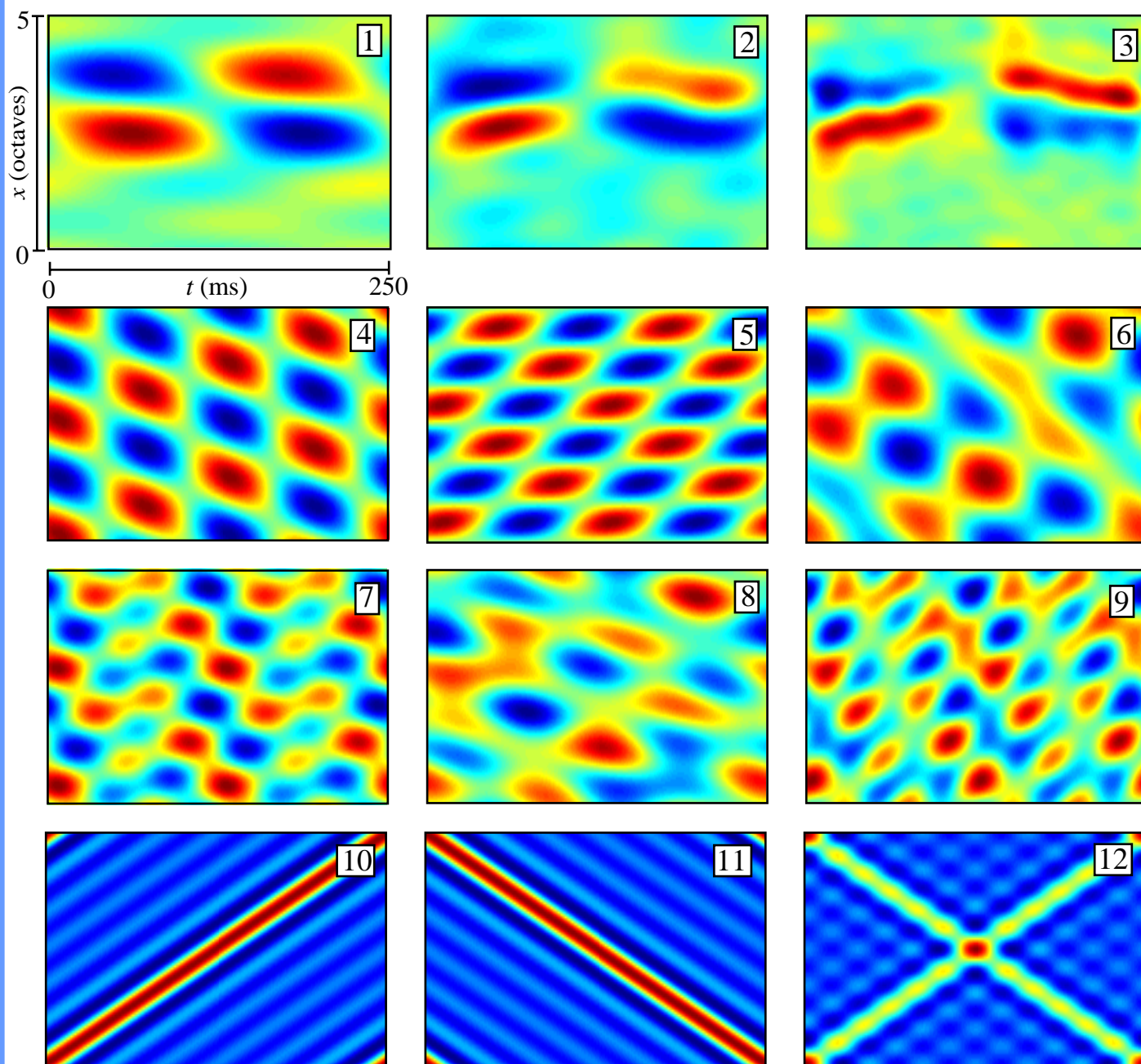
- Introductory Material
- Stimulus Types
- Spectro-Temporal Response Fields
- **Predicting Responses to Novel Stimuli**
- Non-Linearities

# Predicting Responses

- Predictions of Responses to Novel Stimuli

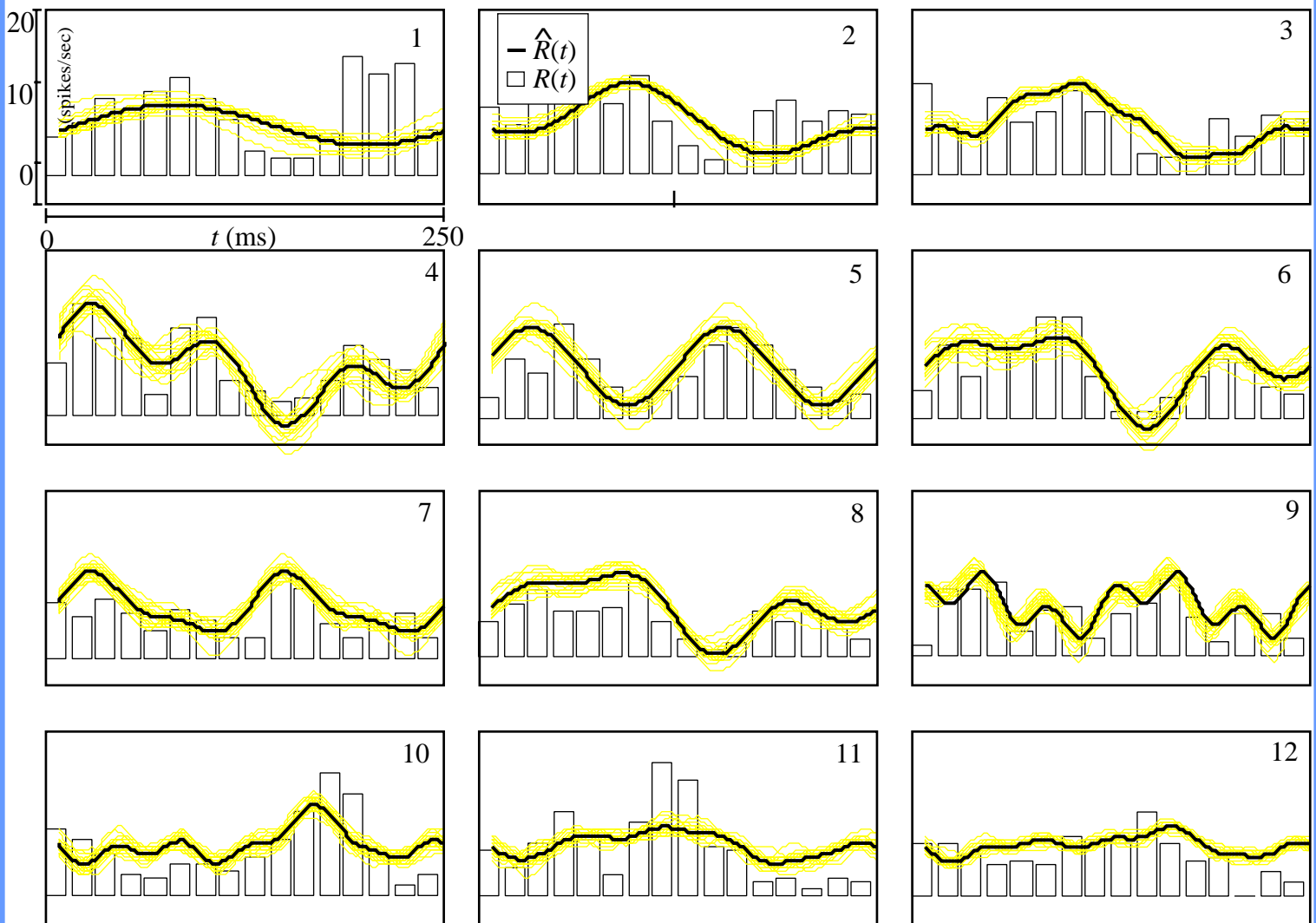
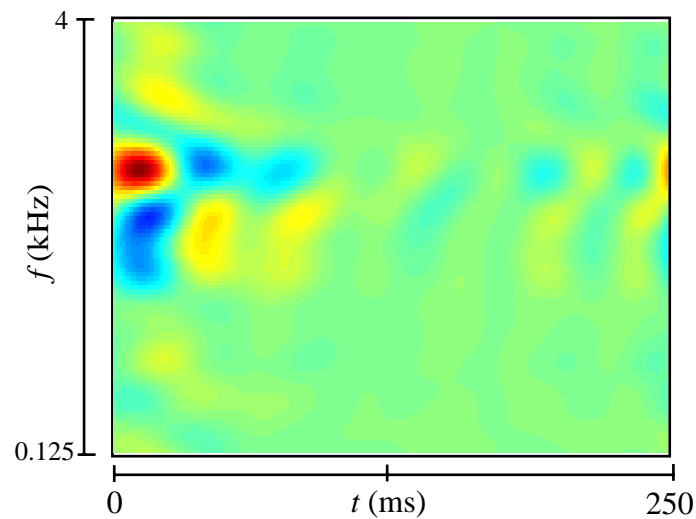
Good test of linearity

## Stimuli Used for Predictions



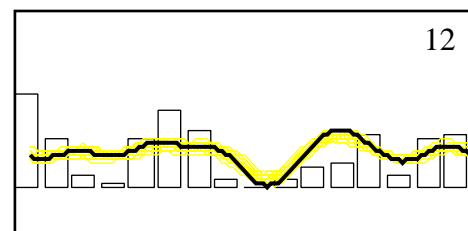
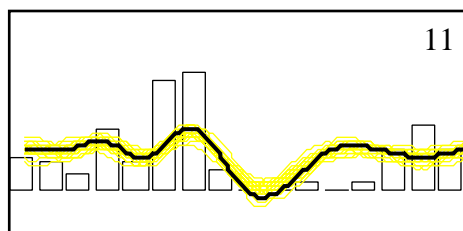
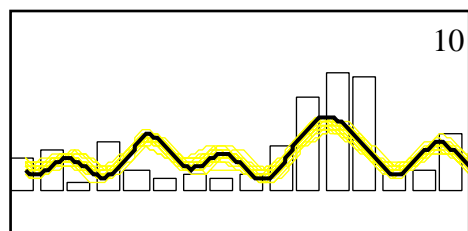
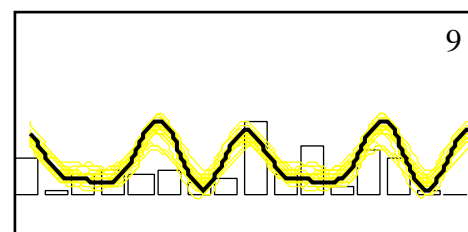
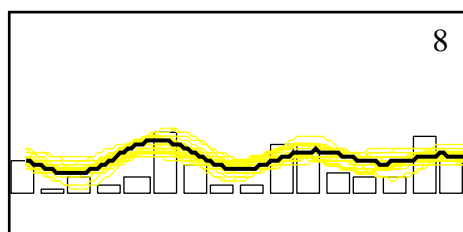
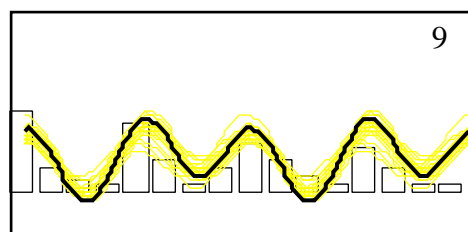
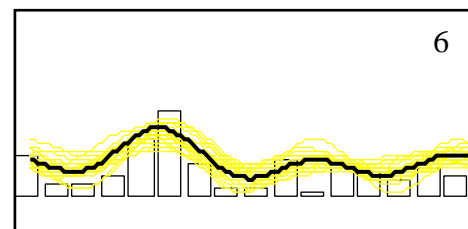
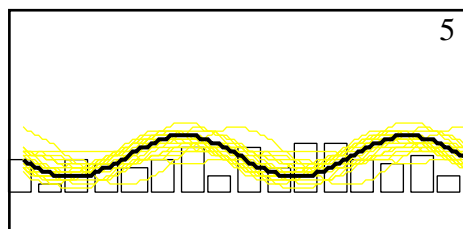
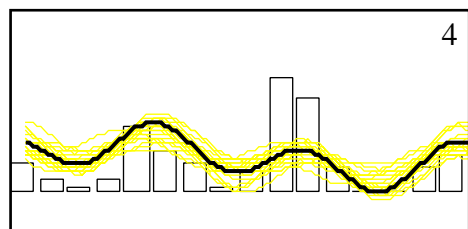
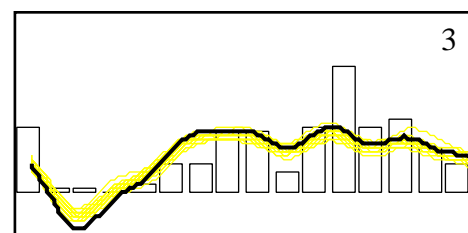
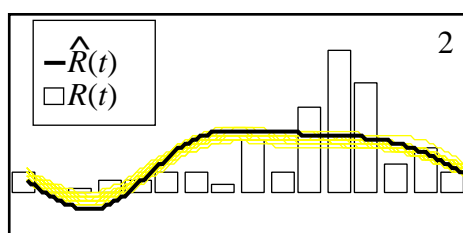
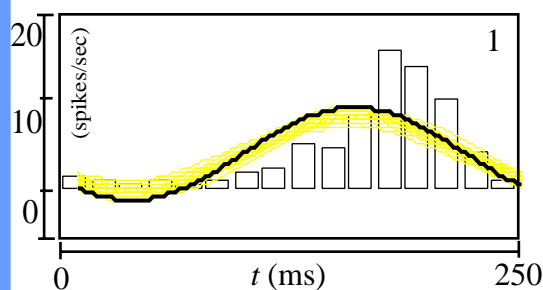
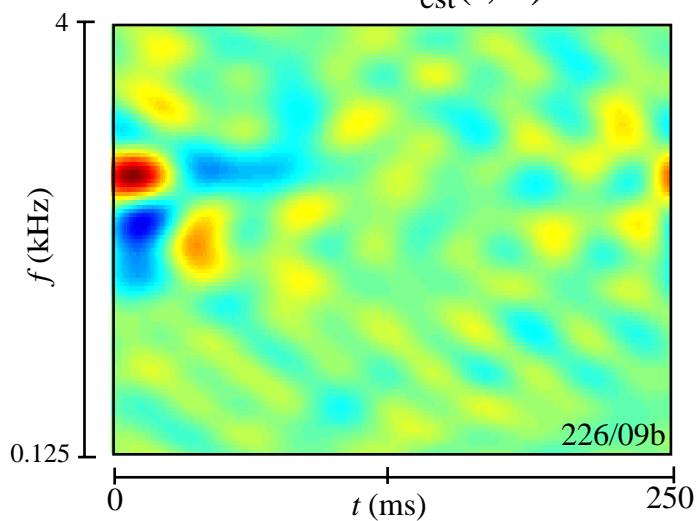
# Predictions and Responses

The STRF estimates often predict the magnitude and dynamics of the response well.



# Predictions and Responses II

$$\text{STRF}_{\text{est}}(t, x)$$

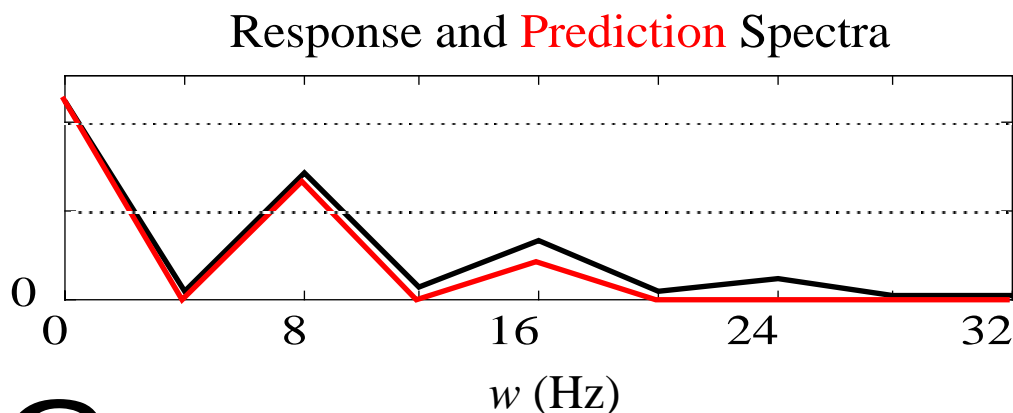
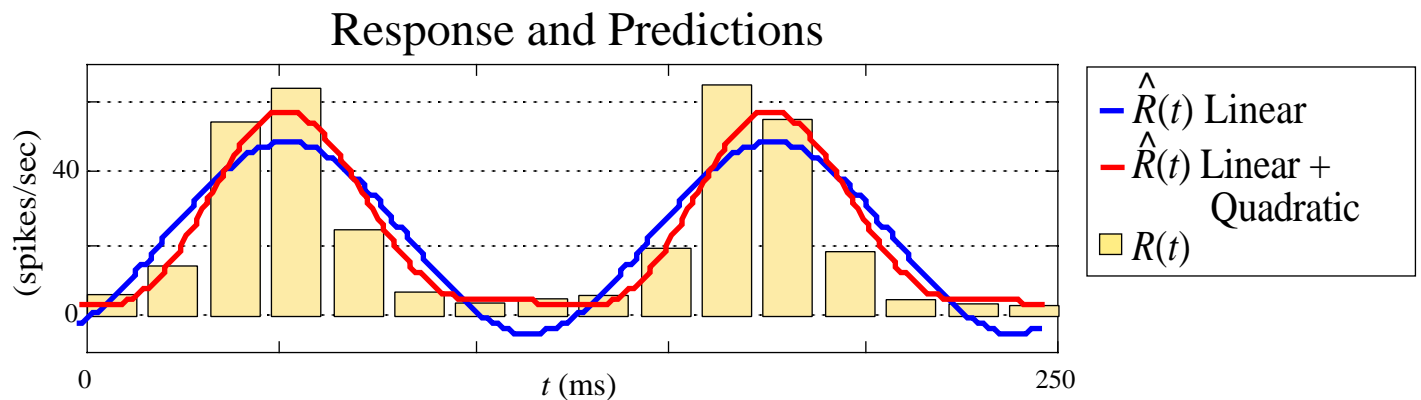
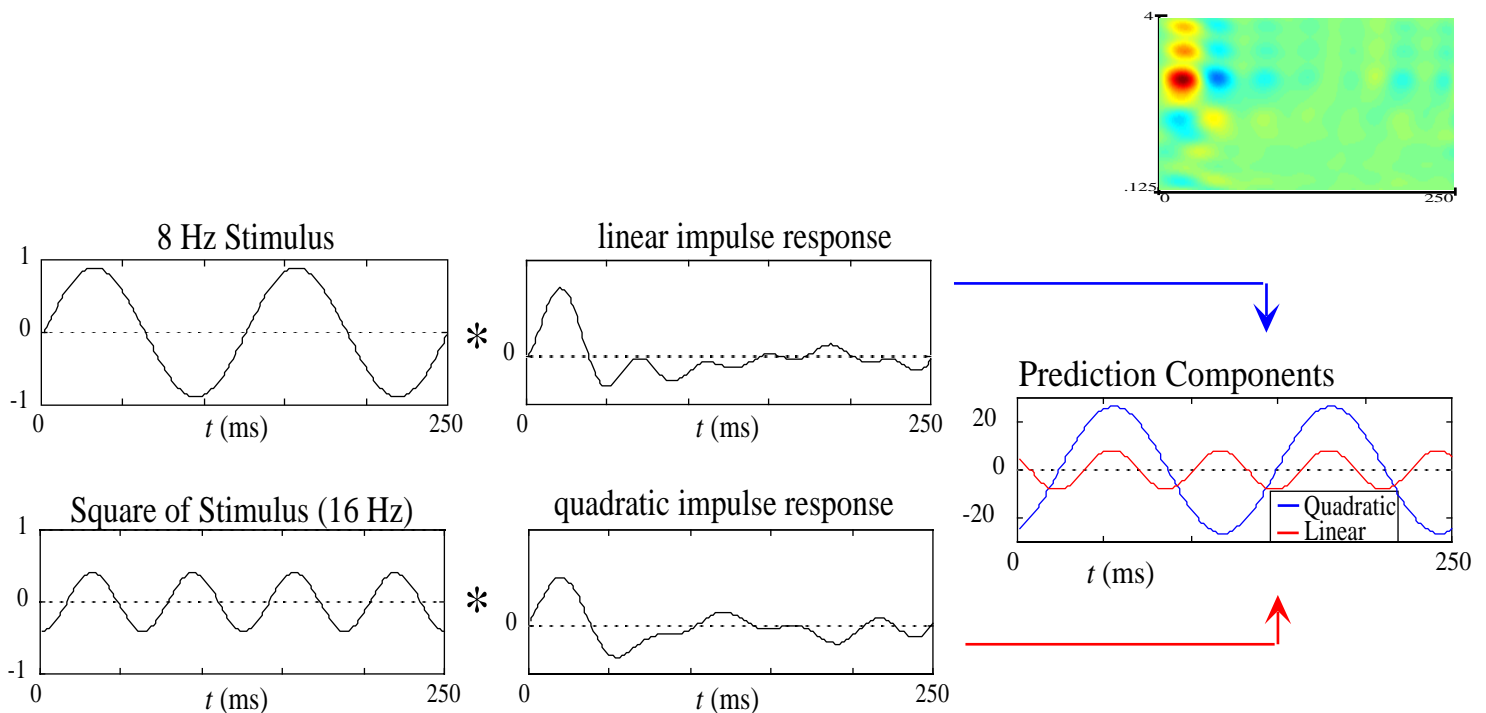




- Introductory Material
- Stimulus Types
- Spectro-Temporal Response Fields
- Predicting Responses to Novel Stimuli
- **Non-Linearities**

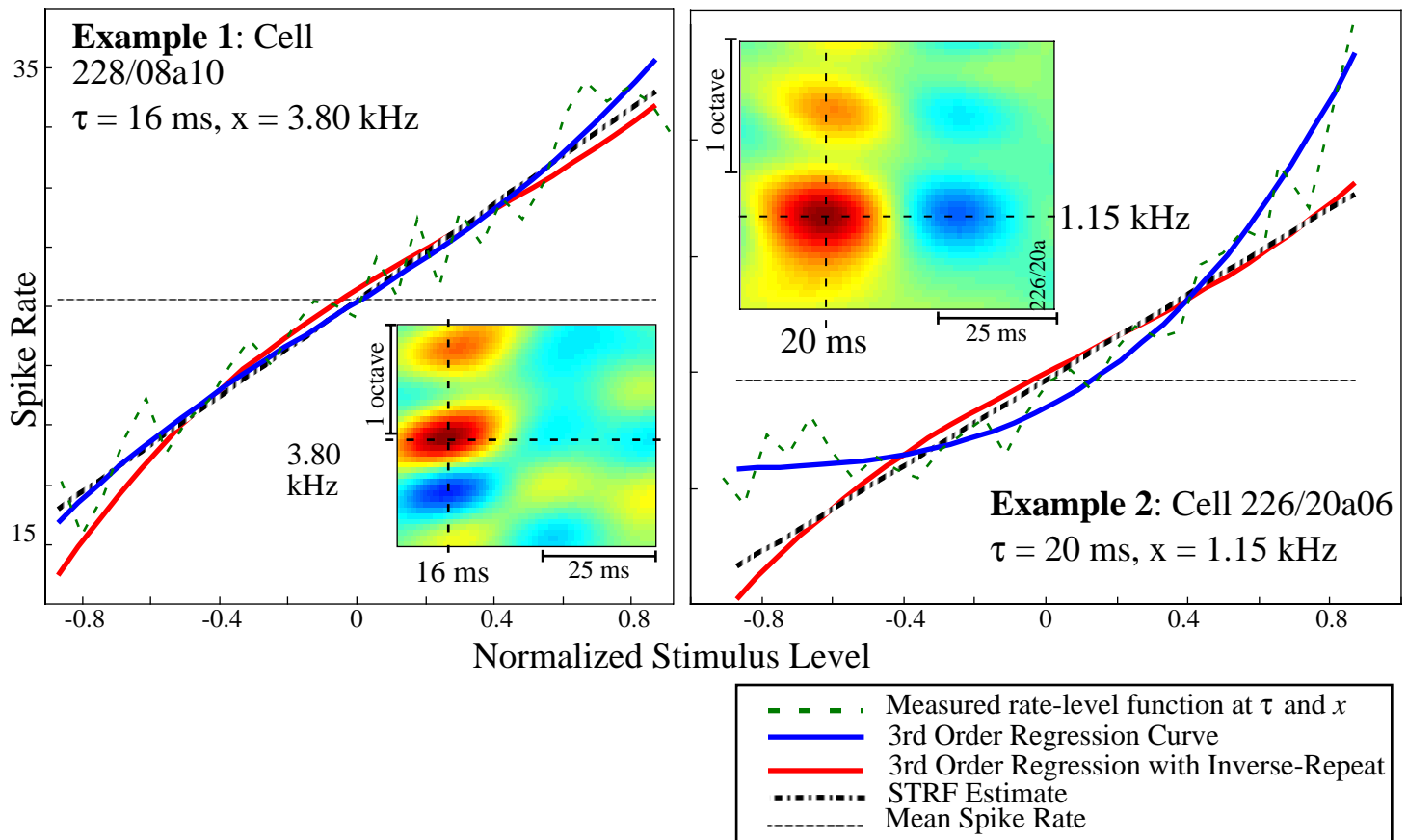
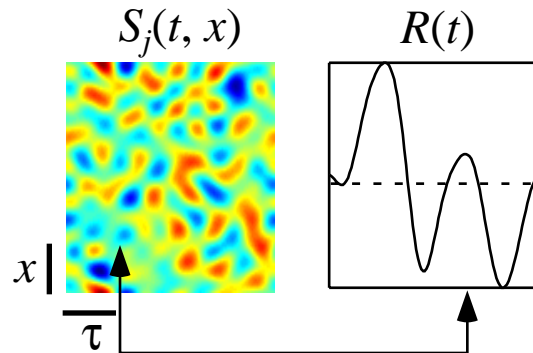
# Non-Linearity—Predictions

- Preliminary results indicate that the non-linear predictions fit the responses more accurately than the linear predictions, although the differences between the two are typically subtle.



# Non-Linearity—Theory

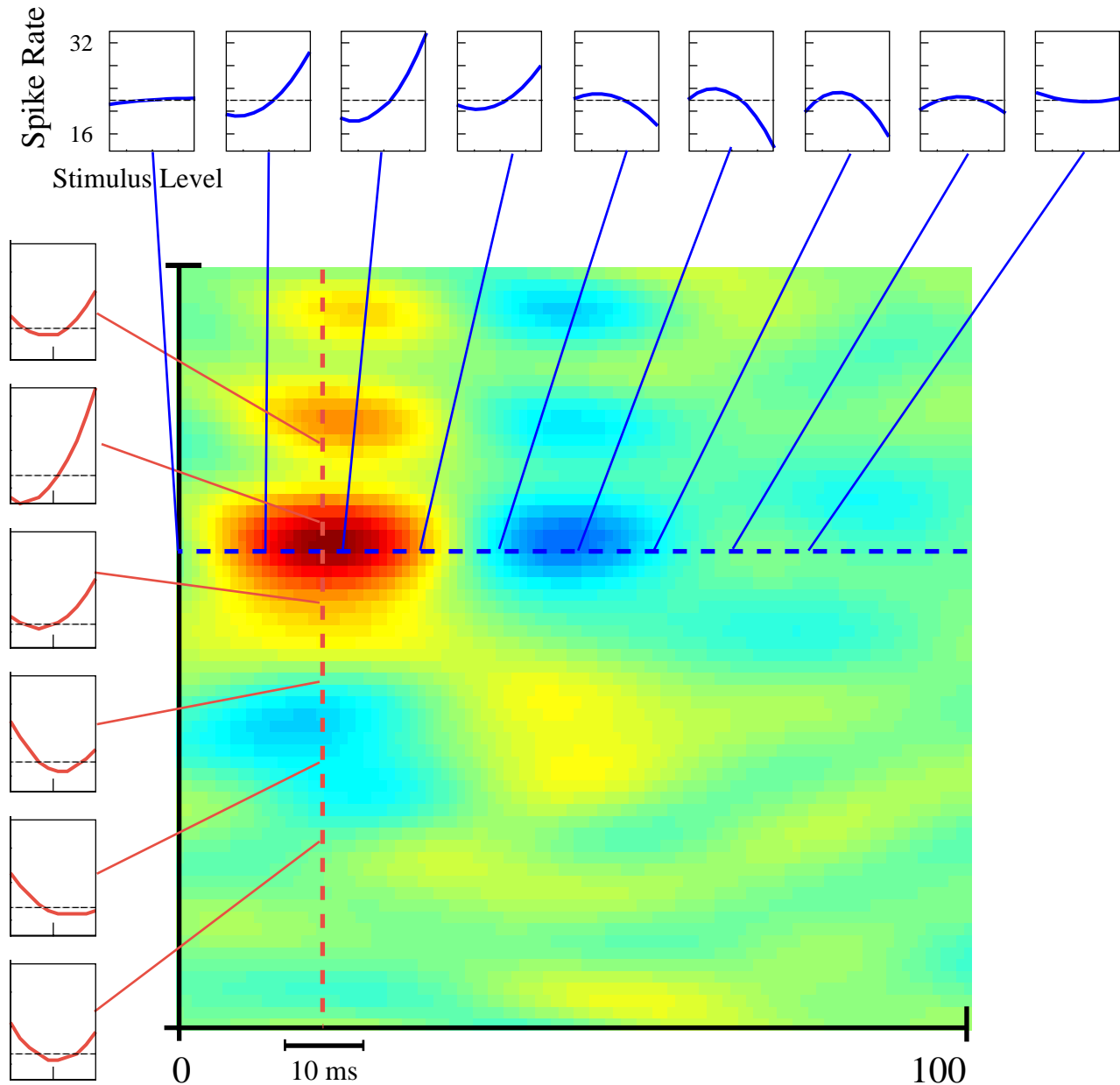
- The value of the STRF at each point  $(\tau, x)$  is the slope of a linear rate-level function:  $R_{\tau,x}(t) = [\text{STRF}(\tau, x)] \cdot S(t - \tau, x)$ .
- Polynomial rate-level curves measured at every  $(\tau, x)$  improve the description. These are potentially non-linear functions.



- Using cubic polynomials, we have shown that either the non-linearities are absent, or they are dominantly second order.
- Subtraction of the response to the inverted envelope gives a nearly linear polynomial fit. This would be expected, for example, from a purely even order (e.g., rectifying non-linearity).

# Spectro-Temporal Rate-Level Functions

Rate-level functions change with  $\tau$  and  $x$ .



# Summary

- **The function of AI**

To encode spectro-temporal features of sounds

spectrally: up to  $\sim 1$  cycles/octave (rarely up to 4 c/o)

temporally:  $\sim 2$  to  $\sim 20$  Hz (rarely up to 100 Hz)

(in ferret)

*plus, of course, to encode other sound features not addressed here*

- **Spectro-Temporal Response Field (STRF)**

- **Descriptor** of spike rate for broadband dynamic stimuli

- **Predictor** of spike train for stimuli of dynamic, spectral modulations of noise

- STRFs agree despite measurement method

- Predictions of responses to novel stimuli

- **Visual Tool** demonstrates spectro-temporal regions of excitation and inhibition

- **Physiological Constraints (e.g. Separability)** constrain possible network dynamics

- **Non-linear corrections/generalizations**  
realizable and work in progress

# Selected References

## **Moving Ripple / STRF / Ripple Transfer Function**

- ❑ Kowalski NA, Depireux DA and Shamma SA, J. Neurophys. 76 (5) (1996) 3503-3523, and 3524-3534.
- ❑ Depireux DA, Simon JZ and Shamma SA, Comments in Theoretical Biology (1999).

## **Spectro-Temporal Correlation Methods**

- ❑ Calhoun BM, Miller RL, Wong JC, and Young ED, 11th International Symposium on Hearing (1997).
- ❑ Eggermont JJ, Hearing Research 66 (1993) 177-201.

## **Singular Value Decomposition**

- ❑ Pilgram B, Schappacher W, Int. J. Bifurcation and Chaos 8 (1999) 571-580.

## **Polynomial Correlation**

- ❑ van Dijk P, Wit HP, Segenhout JM, Hear. Res. 114 (1997) 229-242.
- ❑ de Boer E, Proc. IEEE 67 (1979) 317-318.

## **Related techniques and models**

- ❑ Schreiner CE and Calhoun BM. Auditory Neurosci., 1 (1994) 39-61.
- ❑ Wang K and Shamma SA, IEEE Trans. on Speech and Audio 2(3) (1994) 421-435, and 3(2) (1995) 382-395.
- ❑ Shamma SA, Fleshman JW, Wiser PR and Versnel H, J. Neurophys 69(2) (1993) 367-383.
- ❑ Shamma SA, Versnel H and Kowalski NA, J. Auditory Neuroscience (1) (1995) 233-254, and 255-270, and 271-285.